

**Projet : Analyse de la productivité du cacao en Afrique
subsaharienne : Entre potentiel et fragilité des données**

Par

Gérard De La Paix BAYIHA

(Economiste agricole et Data analyst junior)

gerarddelapaixbayiha@yahoo.fr

Table des matières

1	Objectif	3
2	Limite	3
3	Méthodologie	3
3.1	Présentation de la base de données	3
4	Représentativité des données en fonction de l'origine	4
5	MySQL	5
6	Python	7
6.1	Affichage des informations du dataframe, vérification de l'existence des observations dupliquées	8
7	La pondération de l'origine des données	9
7.1	Test de sensibilité de la production	12
7.2	Test de sensibilité des superficies récoltées	13
7.3	Test de sensibilité des rendements	15
7.4	Conclusion sur les tests de sensibilité	16
8	Evolution comparée des rendements de cacao entre les trois pays	16
9	Evolution comparée des productions de cacao entre les trois pays	16
10	Evolution comparée des superficies de cacao entre les trois pays (2012-2022)	17
11	Conclusion et recommandation	18

Nous avons analysé la productivité du cacao en Afrique subsaharienne à travers un focus sur trois producteurs majeurs : Cameroun, Côte-d'Ivoire et Ghana. Pour ce faire, nous avons obtenu une base de données de FAOSTAT qui couvre la période 2012-2022, soit 10 ans.

1 Objectif

Comparer la productivité du cacao entre les pays d'Afrique centrale et de l'Ouest à partir de la base de données FAOSTAT.

2 Limite

Nous nous sommes uniquement concentrés sur les données provenant de la FAO. Nous n'avons pas fait de croisement avec d'autres sources de bases de données.

3 Méthodologie

Pour manipuler cette base de données, nous allons combiner deux outils : un système de gestion de base de données (MySQL) pour améliorer la structure de la base à travers un Modèle Logique de Données et Python pour faire l'analyse exploratoire de données.

Nous travaillerons d'abord avec les données brutes obtenues de FAOSTAT dans laquelle la colonne Élément regroupe tous les indicateurs. Par la suite, nous mobiliserons MySQL pour mieux gérer la base de données. Enfin, les dataframes obtenus nous permettront de faire des analyses statistiques.

3.1 Présentation de la base de données

Nous avons une base de données de 99 observations et 15 variables. Parmi ces variables, nous avons huit qui sont négligeables et qui ont été supprimées. Nous sommes restés avec 7 variables :

- **Zone** : renvoie aux pays (Cameroun, Côte d'Ivoire, Ghana)
- **Elément** : correspond aux indicateurs (rendement, production, superficie récoltée)
- **Produit** : la culture qui est le cacao
- **Année** : 2012 à 2022
- **Unité** : renvoie aux unités de chacun des indicateurs
- **Description du symbole** : donne l'information sur l'origine des données (chiffre officielle, valeur estimée, valeur imputée, chiffre de source internationale)
- **Note** : donne des détails précis sur la description du symbole. Elle relève que les chiffres de source internationale sont non officiels.

4 Représentativité des données en fonction de l'origine

Pour présenter cela, nous allons nous appuyer sur les données brutes. Il est important d'avoir un aperçu sur la proportion de l'origine des données pour chaque zone. La raison que la colonne **Description du symbole** met en évidence qu'elles proviennent de quatre sources : chiffre officielle, valeur estimée, valeur imputée, chiffre de source internationale. Le résultat de cette information est important sur la crédibilité (degré de confiance et qualité de la comparaison des indicateurs entre les pays) de l'analyse économique qui sera faite.

A travers Python, nous avons croisé les informations relatives à la description du symbole et à la zone dans le but d'obtenir le poids de représentativité des sources d'informations. Nous obtenons ainsi le tableau et la figure ci-dessous.

```
# Croiser Pays et Description
table_quality = df.groupby(["Zone", "Description du Symbole"]).size().reset_index(name="nb_obs")

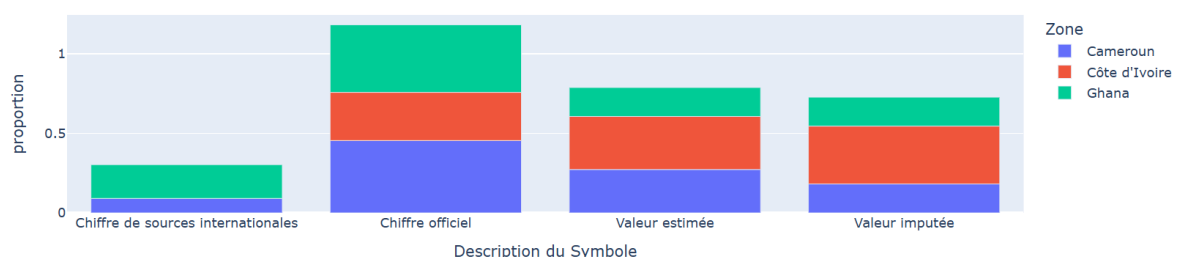
# Normaliser en proportion par pays
table_quality["proportion"] = table_quality.groupby("Zone")["nb_obs"].transform(lambda x: x / x.sum())

print(table_quality)
```

	Zone	Description du Symbole	nb_obs	proportion
0	Cameroun	Chiffre de sources internationales	3	0.090909
1	Cameroun	Chiffre officiel	15	0.454545
2	Cameroun	Valeur estimée	9	0.272727
3	Cameroun	Valeur imputée	6	0.181818
4	Côte d'Ivoire	Chiffre officiel	10	0.303030
5	Côte d'Ivoire	Valeur estimée	11	0.333333
6	Côte d'Ivoire	Valeur imputée	12	0.363636
7	Ghana	Chiffre de sources internationales	7	0.212121
8	Ghana	Chiffre officiel	14	0.424242
9	Ghana	Valeur estimée	6	0.181818
10	Ghana	Valeur imputée	6	0.181818

```
fig = px.bar(table_quality, x="Description du Symbole", y="proportion",
             hover_data=["Zone"],
             color="Zone",
             title="Représentativité de l'origine des données par pays")
fig.show()
```

Représentativité de l'origine des données par pays



De ces informations, nous relevons que le Cameroun et le Ghana avec respectivement 45 et 42 % ont une fiabilité des données du fait caractère officiel que la Côte d'Ivoire (10%). De plus, nous constatons que la Côte d'Ivoire a une forte proportion (environ 33%) de données imputées que les deux autres pays (environ 18% chacun). Enfin, il en ressort que les institutions internationales fournissent d'importantes informations pour le Ghana (environ 21%). Cet accompagnement est très faible au Cameroun (environ 09%) et inexistant en Côte d'Ivoire.

Ainsi donc, nous constatons une hétérogénéité au sein de l'origine des sources d'informations dans cette base. Pour faire une interprétation rigoureuse, il sera important de pondérer les résultats issus des comparaisons en fonction de la qualité des données.

5 MySQL

Nous travaillons d'abord avec MySQL pour segmenter la colonne « indicateur ». Nous souhaitons avoir trois vues relatives à chacun des indicateurs. Pour ce faire, la première étape vise à :

- Création des tables (Modèle Conceptuel Données) et construction du Modèle Logique de Données

Nous avons créé 04 tables : Valeur, Pays, Année, Culture, Indicateur

```
-- Création de la table pays

create table Pays(
  IDPays int primary key auto_increment,
  Nom_Pays varchar (250)
);

-- Création de la table indicateurs

create table Indicateurs(
  IDIndicateurs int primary key auto_increment,
  Nom_Indicateur varchar(100),
  Unité_Indicateur varchar(200)
);

select*
from Indicateurs;

-- Création de la table Culture

create table Culture (
  IDCulture int primary key auto_increment,
  Nom_Culture varchar(250)
);

-- Création de la table Année

create table Année (
  IDAnnée int primary key auto_increment,
  Année int);
```

```

-- Création de la table Valeur
create table Valeur(
  IDValeur int primary key auto_increment,
  Valeur int,
  IDIndicateurs int,
  IDCulture int,
  IDAnnée int,
  IDPays int,
  FOREIGN KEY(IDIndicateurs) references Indicateurs(IDIndicateurs),
  FOREIGN KEY(IDCulture) references Culture(IDCulture),
  FOREIGN KEY(IDAnnée) references Année(IDAnnée),
  FOREIGN KEY(IDPays) references Pays(IDPays)
);

```

- Modèle logique de données

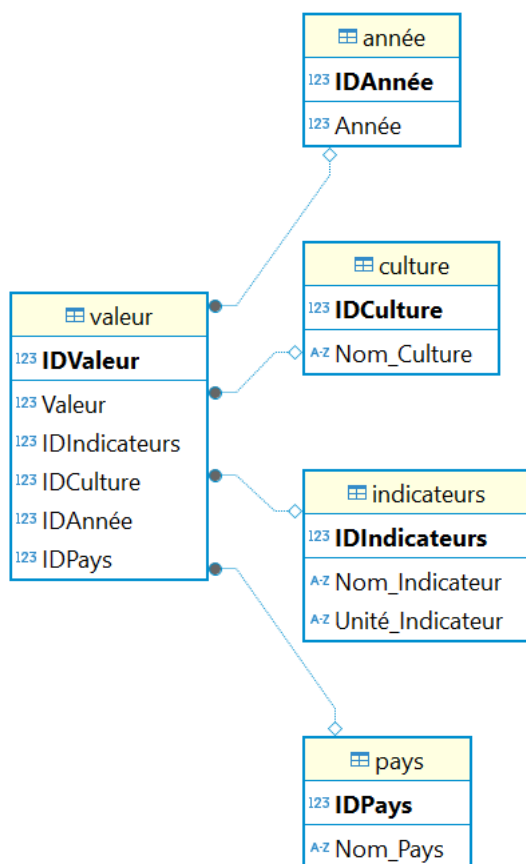


Figure : Modèle Logique de données

- Nous importons les données de la base de FAOSTAT dans chacune des tables.
- On crée des vues pour chaque indicateur

```

-- On créé une vue uniquement pour les productions par pays, culture et par année

create view Production as
select Nom_Pays as Pays, Nom_Culture as Culture, Année, Nom_Indicateur, Valeur as Production,
from valeur
join Pays using(IDPays)
join Culture using (IDCulture)
join Année using (IDAnnée)
join Indicateurs using (IDIndicateurs)
where Nom_Indicateur = "Production" and Unité_Indicateur = "tonnes";

select*
from Production;

-- On créé une vue uniquement pour les rendements par pays, culture et par année

create view Rendement as
select Nom_Pays as Pays, Nom_Culture as Culture, Année, Nom_Indicateur, Valeur as Rendement,
from valeur
join Pays using(IDPays)
join Culture using (IDCulture)
join Année using (IDAnnée)
join Indicateurs using (IDIndicateurs)
where Nom_Indicateur = "Rendement" and Unité_Indicateur = "kg/ha";

select*
from Rendement;

-- On créé une vue uniquement pour les superficies récoltées par pays, culture et par année

create view Superficie_Récoltée as
select Nom_Pays as Pays, Nom_Culture as Culture, Année, Nom_Indicateur, Valeur as Superficie_R
from valeur
join Pays using(IDPays)
join Culture using (IDCulture)
join Année using (IDAnnée)
join Indicateurs using (IDIndicateurs)
where Nom_Indicateur = "Superficie récoltée" and Unité_Indicateur = "ha";

select*
from Superficie_Récoltée;

```

- On exporte les trois tables (production, rendement, superficie_récoltée) sous format csv. Après cela, on nettoie les tables des imperfections (noms mal écrits)

6 Python

Nous avons importé les trois tables générées dans MySQL.

Fusion des trois tables (rendement, production, superficie récoltée)

- On fusionne les trois tables importées dans le but d'avoir une base de données avec des colonnes visibles concernant chaque indicateur, leur valeur, leur unité par pays et par année.

```
# Fusion des trois dataframe (rendement, superficie récoltée, production)
Fusion = (
    rendement.merge(superficie, on =['Pays','Culture'])
    .merge(production,on =['Pays','Culture']))

Fusion.head() # Afficha des cinq premières Lignes
```

	Pays	Culture	Année_x	Nom_Indicateur_x	Rendement	Unité_Indicateur_x	Année_y	Nom_Indicateur_y	Superficie_Récoltée	Unité_Indicateur_y	Année	Nom_In
0	Cameroun	Cacao, fèves	2012	Rendement	401	kg/ha	2012	Superficie récoltée	670000	ha	2012	Pr
1	Cameroun	Cacao, fèves	2012	Rendement	401	kg/ha	2012	Superficie récoltée	670000	ha	2012	Pr
2	Cameroun	Cacao, fèves	2012	Rendement	401	kg/ha	2012	Superficie récoltée	670000	ha	2013	Pr
3	Cameroun	Cacao, fèves	2012	Rendement	401	kg/ha	2012	Superficie récoltée	670000	ha	2013	Pr
4	Cameroun	Cacao, fèves	2012	Rendement	401	kg/ha	2012	Superficie récoltée	670000	ha	2014	Pr

6.1 Affichage des informations du dataframe, vérification de l'existence des observations dupliquées

```
Fusion.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5324 entries, 0 to 5323
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Pays                   5324 non-null   object
1   Culture                5324 non-null   object
2   Année_x               5324 non-null   int64
3   Nom_Indicateur_x      5324 non-null   object
4   Rendement             5324 non-null   int64
5   Unité_Indicateur_x    5324 non-null   object
6   Année_y               5324 non-null   int64
7   Nom_Indicateur_y      5324 non-null   object
8   Superficie_Récoltée   5324 non-null   int64
9   Unité_Indicateur_y    5324 non-null   object
10  Année                 5324 non-null   int64
11  Nom_Indicateur         5324 non-null   object
12  Production             5324 non-null   int64
13  Unité_Indicateur       5324 non-null   object
dtypes: int64(6), object(8)
memory usage: 582.4+ KB
```

Il s'agit d'avoir accès à différentes informations :

- Dimension : 5323 observations et 14 colonnes
- Types de variable : 06 variables quantitatives et 08 qualitatives
- Noms des variables : confère figure
- Données manquantes : Aucune


```
duplicate = Fusion[Fusion.duplicated()]
print("Les données dupliquées sont :")
print(duplicate)
```

Les données dupliquées sont :
Empty DataFrame
Columns: [Pays, Culture, Année_x, Nom_Indicateur_x, Rendement, Unité_Indicateur_x, Année_y, Nom_Indicateur_y, Superficie_Récoltée, Unité_Indicateur_y, Année, Nom_Indicateur, Production, Unité_Indicateur]
Index: []

Au regard de la figure ci-dessus, nous relevons qu'il n'y a pas de données dupliquées.

7 La pondération de l'origine des données

Comme nous l'avons mentionné dans la section 4, du fait de l'hétérogénéité des sources d'information, il est important de procéder à une pondération de chacune de ces sources avant d'analyser les indicateurs.

Dans la littérature, il n'existe pas une méthodologie claire concernant la pondération des indicateurs. Une étude de l'Union Européenne¹(2005) met en évidence qu'il y a plusieurs démarches qui peuvent être envisagées :

- Accorder une pondération plus élevée en fonction de l'importance de l'indicateur selon le chercheur
- Utiliser des pondérations dérivées de l'analyse en composantes principales dans l'optique surmonter les problèmes de double comptage lorsque deux ou plusieurs indicateurs mesurent partiellement le même comportement
- Utiliser des pondérations en se basant sur la corrélation entre les indicateurs
- Réaliser des ateliers au cours desquels les spécialistes s'entendent pour accorder un poids à chaque indicateur.

A cet effet, nous avons opté pour la première option toutefois, en adoptant une approche raisonnée. La figure ci-dessous rappelle la proportion de chaque source de données dans la base.

¹ Nado M., Saisana M., Saltelli A., Tarantola S. (2005). Tools for composite indicators building. European Commission : pp. 1-134

```
# Croiser Pays et Description
table_quality = df.groupby(["Zone", "Description du Symbole"]).size().reset_index(name="nb_obs")

# Normaliser en proportion par pays
table_quality["proportion"] = table_quality.groupby("Zone")["nb_obs"].transform(lambda x: x / x.sum())

print(table_quality)
```

	Zone	Description du Symbole	nb_obs	proportion
0	Cameroun	Chiffre de sources internationales	3	0.090909
1	Cameroun	Chiffre officiel	15	0.454545
2	Cameroun	Valeur estimée	9	0.272727
3	Cameroun	Valeur imputée	6	0.181818
4	Côte d'Ivoire	Chiffre officiel	10	0.303030
5	Côte d'Ivoire	Valeur estimée	11	0.333333
6	Côte d'Ivoire	Valeur imputée	12	0.363636
7	Ghana	Chiffre de sources internationales	7	0.212121
8	Ghana	Chiffre officiel	14	0.424242
9	Ghana	Valeur estimée	6	0.181818
10	Ghana	Valeur imputée	6	0.181818

Nous avons pour chaque source de données, pris la moyenne pondérée de la proportion que chacune d'elle représente dans la base. Ainsi, nous avons le tableau suivant :

Tableau : pondération de l'origine des sources de données dans la base de FAOSTAT

Origine	Calcul	Poids
Chiffre officiel (A)	$(0.45 + 0.3 + 0.42)/3$	0.4
Valeur estimée (E)	$(0.27 + 0.33 + 0.18)/3$	0.28
Valeur imputée (I)	$(0.18+0.36+0.18)/3$	0.24
Internationale (X)	$(0.09 + 0.21)/2$	0.15

Source : Auteur

Après chaque représentation, nous ferons un **test de sensibilité univariée (Briggs et al., 1994)** par variation des paramètres. En d'autres termes, cette méthode consiste à faire varier de (+/- 0.10) un paramètre à la fois tout en maintenant les autres constants. Il mesure la robustesse d'un indicateur quand on change légèrement les hypothèses à travers trois scénarios :

- Celui de base (il est en lien avec la pondération proposée)
- Celui bas (-0.10)
- Celui haut (+ 0.10)

Dans cette étude, les hypothèses sont liées aux pondérations appliquées selon l'origine des données (officielles, estimées, imputées, sources internationales).

Sur le plan visuel, ces scénarios sont représentés sur une droite par :

- Celui de base par un point
- Celui bas et celui haut, par un segment vertical (intervalle de sensibilité). Lorsque cet intervalle est petit, cela veut dire que le test est robuste, c'est-à-dire qu'il est peu sensible aux changements de pondération des données. A l'opposé, il l'est plus sensible.

Le code script sur python utilisé pour faire la table du test de sensibilité et le graphique associé est le suivant :

Pondération et test de sensibilité des indicateurs : rendement, superficie récoltée et production

```
] : # 1. Définition des poids de qualité

# Pondérations de base issue de la démarche expliquée dans le fichier word
# A : chiffre officiel ; E : valeur estimée ; I : valeur imputée ; X : sources internationales
# La méthodologie pour obtenir les valeurs de ces poids (weights_base) est donnée dans le fichier word

weights_base = {"A": 0.4, "E": 0.28, "I": 0.24, "X": 0.15}

# Scénario haut (+0.10) et bas (-0.10)
weights_high = {k: min(1.0, v + 0.10) for k, v in weights_base.items()}
weights_low = {k: max(0.0, v - 0.10) for k, v in weights_base.items()}

# 2. Fonction d'application des poids

def apply_weights(df, weight_map, suffix="_pond"):
    """Applique les pondérations aux indicateurs"""
    df = df.copy()
    df["poids"] = df["Lettre_Flag"].map(weight_map).fillna(0.5) # défaut = 0.5 comme Lettre_Flag inconnu
    for col in ["Production", "Rendement", "Superficie_Récoltée"]:
        df[col + suffix] = df[col] * df["poids"]
    return df

# 3. Appliquer les 3 scénarios

df_base = apply_weights(Fusion, weights_base, "_pond")
df_low = apply_weights(Fusion, weights_low, "_pond")
df_high = apply_weights(Fusion, weights_high, "_pond")

# Agrégats par pays
agg_base = df_base.groupby("Pays", as_index=False)[indicator + "_pond"].sum().rename(columns={indicator + "_pond": "Baseline"})
agg_low = df_low.groupby("Pays", as_index=False)[indicator + "_pond"].sum().rename(columns={indicator + "_pond": "Min"})
agg_high = df_high.groupby("Pays", as_index=False)[indicator + "_pond"].sum().rename(columns={indicator + "_pond": "Max"})

merged = agg_base.merge(agg_low, on="Pays").merge(agg_high, on="Pays")

# Affichage du tableau
print(f"\n=== Test de sensibilité : {indicator} ===")
print(merged)

# --- Graphique ---
plt.figure(figsize=(8,5))
for i, row in merged.iterrows():
    # Intervalle min-max (trait vertical)
    plt.plot([i, i], [row["Min"], row["Max"]], color="gray", linewidth=2)
    # Valeur baseline (point bleu)
    plt.scatter([i], [row["Baseline"]], color="blue", s=80, zorder=3)
```

```

plt.xticks(range(len(merged["Pays"])), merged["Pays"], rotation=0)
plt.ylabel(indicator)
plt.title(f"Test de sensibilité aux pondérations ( $\pm 0,10$ ) - {indicator}")
plt.grid(True, linestyle="--", alpha=0.5)
plt.tight_layout()
plt.show()

return merged

# Lancement de l'analyse des trois indicateurs

results_prod = sensitivity_analysis(df_base, df_low, df_high, "Production")
results_rend = sensitivity_analysis(df_base, df_low, df_high, "Rendement")
results_sup = sensitivity_analysis(df_base, df_low, df_high, "Superficie_Récoltée")

```

7.1 Test de sensibilité de la production

Le résultat du test de sensibilité de la production pour les trois pays est analysé à deux niveaux. Le premier porte sur le tableau des indicateurs. Il met en évidence que la Côte-d'Ivoire a le plus haut niveau de production pondérée des trois pays avec environ 8 200 000 tonnes. Lorsque l'on fait varier de (+/- 0.10) ce chiffre, il se situe autour de l'intervalle d'environ 6 000 000 et 10 000 000 tonnes. Le Ghana vient en deuxième position puis arrive le Cameroun.

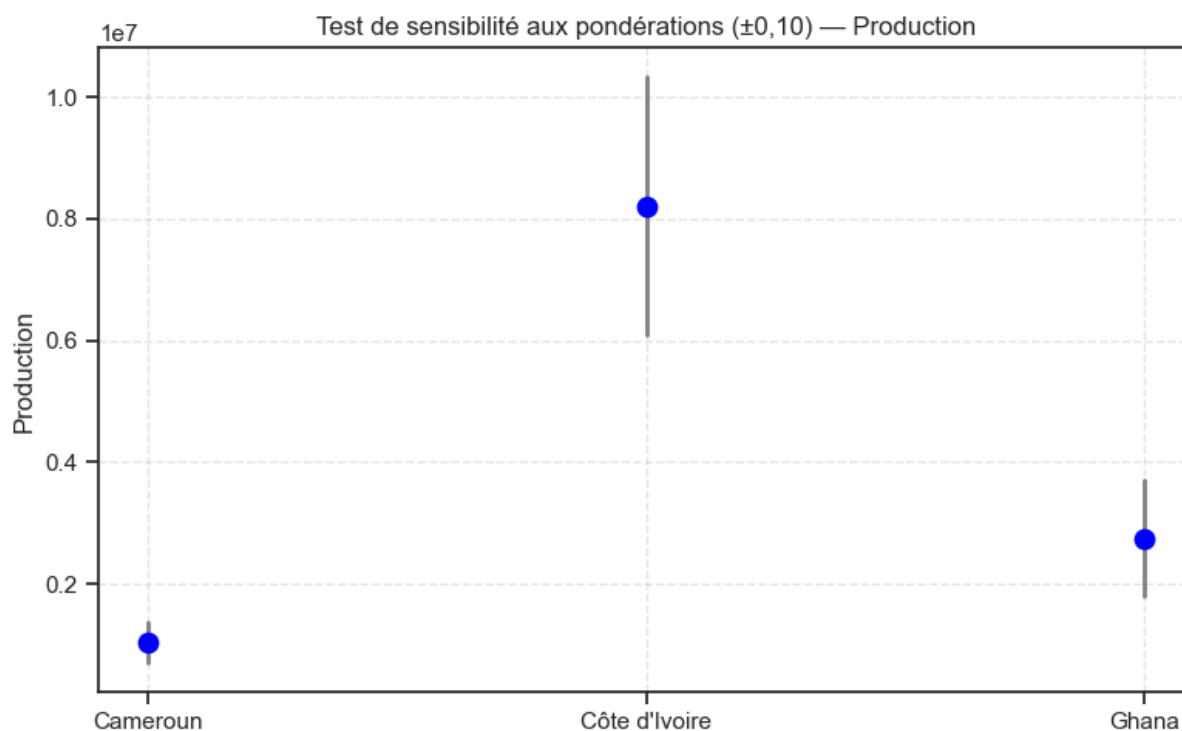
```

=== Test de sensibilité : Production ===

```

	Pays	Baseline	Min	Max
0	Cameroun	1042785.52	715382.52	1370188.52
1	Côte d'Ivoire	8206888.88	6089655.58	10324122.18
2	Ghana	2746013.60	1795277.70	3696749.50

De plus, la lecture graphique de la figure ci-dessous révèle que les segments verticaux sont très petits. Cet aspect confirme la robustesse de ces résultats au niveau de la position hiérarchique de chaque pays.



En conclusion, nous relevons que les intervalles ne se chevauchent pas que ce soit à travers les intervalles qu'au niveau des segments. Même dans le scénario le plus défavorable (pondération donnant plus de poids aux données incertaines), la Côte d'Ivoire reste largement en tête. Nous pouvons donc confirmer que, malgré qu'il y a une incertitude élevée dans la qualité de données, pour quel que soit le pays et le poids donné à l'indicateur de production, la Côte – d'Ivoire aura toujours le niveau de production le plus élevé, suivi du Ghana et enfin le Cameroun. Cela résultat est conforme avec la littérature (Fountain & Huetz-Adams, 2022 ; ICCO, 2023) qui montrent la Côte d'Ivoire comme leader incontesté de la production régionale de cacao.

7.2 Test de sensibilité des superficies récoltées

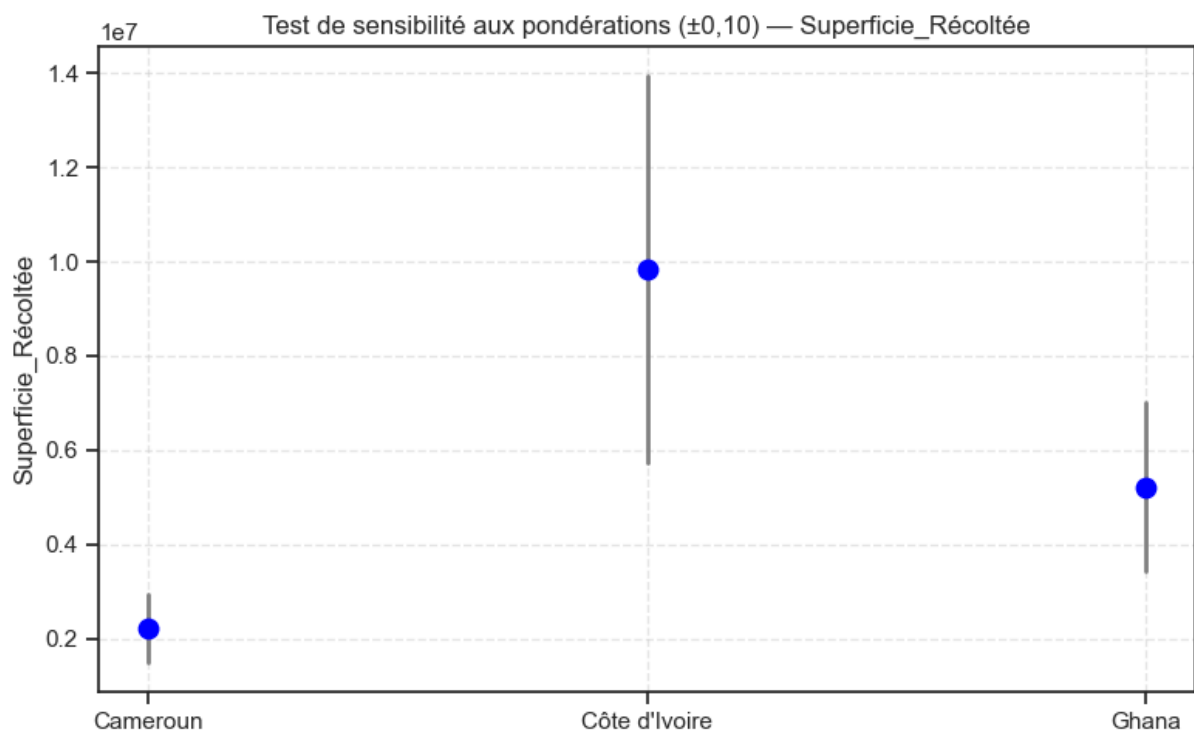
Comme précédemment, le résultat est analysé à deux niveaux. Concernant le tableau du test de sensibilité, il présente que la Côte – d'Ivoire utilise de plus grandes superficies (environ 9 800 000 hectares) que le Ghana (environ 5 200 000 ha) et enfin le Cameroun (2 200 000 ha). Par ailleurs, les amplitudes des intervalles sont importantes et laisse donc paraître un niveau d'incertitude non négligeable de la qualité des données.

```

=== Test de sensibilité : Superficie_Récoltée ===
      Pays      Baseline      Min      Max
0    Cameroun  2217869.12  1494909.52  2940828.72
1 Côte d'Ivoire  9838678.56  5739229.16  13938127.96
2      Ghana   5212262.75  3415220.75   7009304.75

```

Entre autres, la figure ci-dessous vient appuyer de manière visuelle le tableau de test de sensibilité. Nous observons la longueur des intervalles de sensibilité sont beaucoup plus grand que pour les rendements ce qui confirme un niveau d'incertitude important. Cependant, ils ne se chevauchent pas.



En conclusion, malgré le niveau d'incertitude important, les résultats sont robustes du fait qu'il sur le classement entre ces trois pays en matière d'usage des superficies.

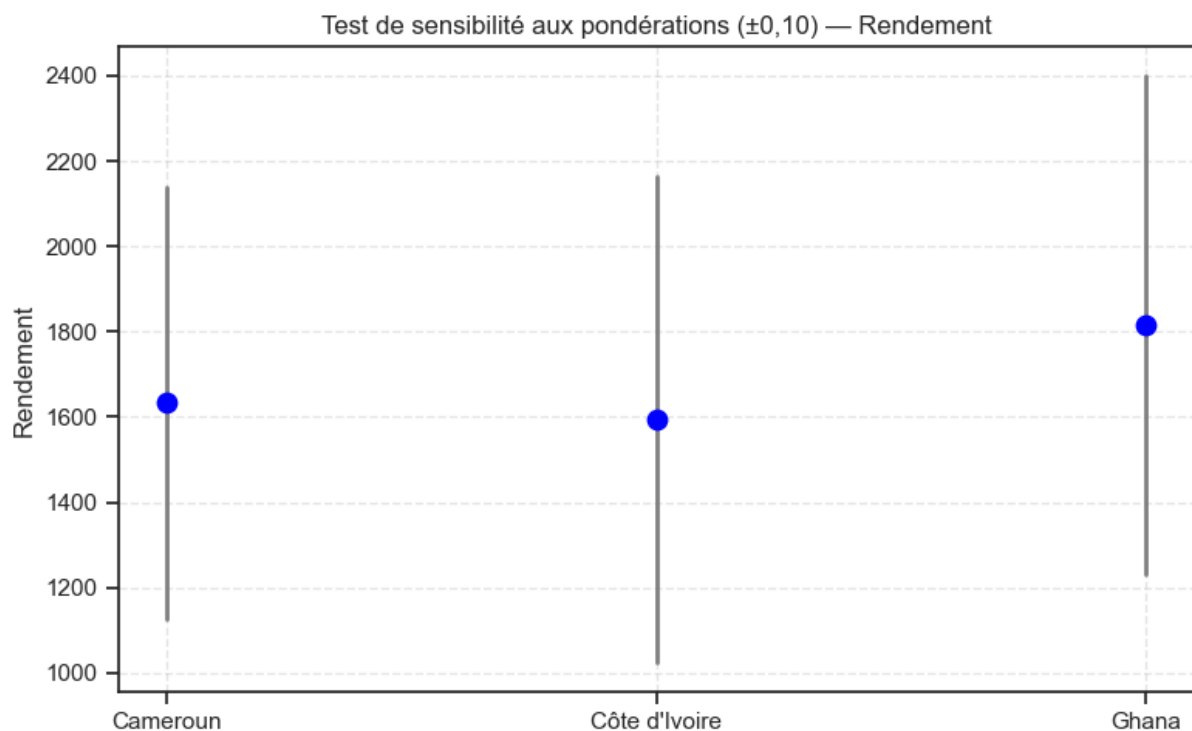
7.3 Test de sensibilité des rendements

Toujours dans le même ordre d'idées d'analyse que précédemment, les résultats du tableau montrent que le Ghana a un rendement de base (Environ 1800 kg/ha) que les deux autres pays. Mais l'écart entre ces rendements de base n'est pas très grand. A travers les intervalles de sensibilité, nous relevons qu'il y a chevauchement. En fonction des pondérations appliquées, le Cameroun peut apparaître supérieur au Ghana, voir à la Côte d'Ivoire, ou l'inverse. Cela rend les comparaisons de rendement moins robustes : la hiérarchie entre pays n'est pas stable.

=== Test de sensibilité : Rendement ===

	Pays	Baseline	Min	Max
0	Cameroun	1633.08	1128.18	2137.98
1	Côte d'Ivoire	1594.88	1025.28	2164.48
2	Ghana	1815.90	1233.30	2398.50

De plus, la lecture du graphique confirme la faible robustesse au niveau de la hiérarchisation des pays. Les segments sont très grands et se chevauchent.



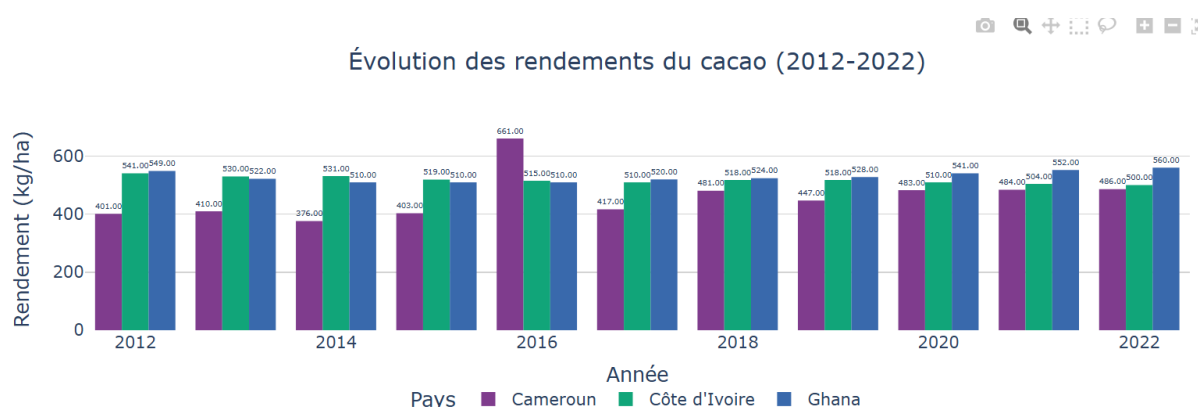
En conclusion, les résultats mettent en évidence un niveau d'incertitude élevée au niveau de la qualité des données et une faible robustesse concernant la hiérarchisation des pays. Ce résultat est conforme avec la littérature (ICCO, 2023), qui souligne les difficultés de comparer des rendements issus de bases hétérogènes.

7.4 Conclusion sur les tests de sensibilité

Les tests de sensibilité montrent que les trois indicateurs présentent une incertitude sur la qualité des données. Cependant, pour la hiérarchisation des pays en matière production et la superficie récoltée du cacao, l'ordre hiérarchique est robuste. Donc, on peut analyser avec une certaine fiabilité l'évolution des diagrammes à barre de ces deux indicateurs tout en attirant les décideurs d'améliorer la production de données statistiques. Par contre, au niveau du rendement, la robustesse concernant la hiérarchie n'est pas validée. Il est donc risqué sur la base des données de la FAOSTAT, de faire une analyse rigoureuse pour les décideurs en matière de stratégies de développement du cacao uniquement sur les rendements.

8 Evolution comparée des rendements de cacao entre les trois pays

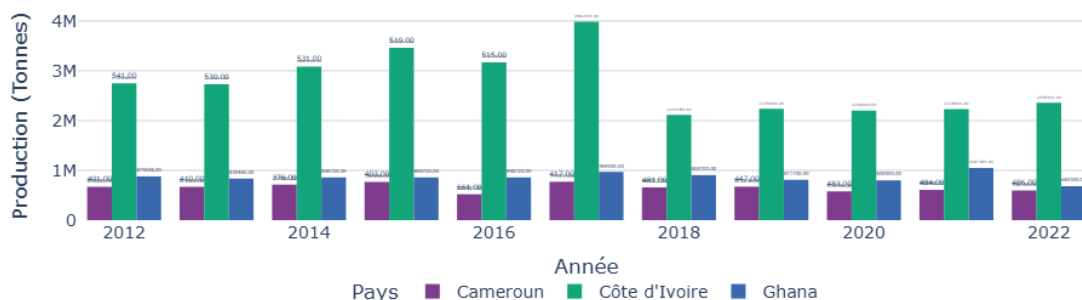
La figure ci-dessous présente l'évolution des rendements. Concernant aux deux indicateurs précédents, il n'y a pas une domination significative d'un pays par rapport à l'autre. Les niveaux moyens apparaissent proches entre les trois pays : 1 633 kg/ha au Cameroun, 1 595 kg/ha en Côte d'Ivoire et 1 816 kg/ha au Ghana. Cependant, les intervalles d'incertitude se chevauchent largement, ce qui remet en cause toute hiérarchie stricte. Le Cameroun, par exemple, pourrait se situer en dessous de la Côte d'Ivoire dans certains scénarios, mais aussi la dépasser dans d'autres. Cette instabilité, confirmée par le test de sensibilité, révèle que les comparaisons de rendement sont sensibles à la qualité et à l'origine des données. Contrairement à la production, la lecture des performances en termes de rendement doit donc rester prudente.



9 Evolution comparée des productions de cacao entre les trois pays

La figure ci-dessous présente l'évolution des productions de cacao entre 2012 et 2022 pour le Cameroun, la Côte d'Ivoire et le Ghana. Elle révèle la domination incontestable de la Côte d'Ivoire, suivie du Ghana et enfin du Cameroun. Le test de sensibilité confirme la robustesse de cette hiérarchie. De plus, même dans les scénarios les plus défavorables, la Côte d'Ivoire reste largement en tête, avec des volumes très supérieurs à ceux du Ghana et du Cameroun. Cette stabilité traduit l'effet d'échelle et l'organisation de la filière ivoirienne, où les volumes produits ne dépendent pas uniquement des hypothèses de pondération des données, mais reposent sur une trajectoire structurelle de croissance.

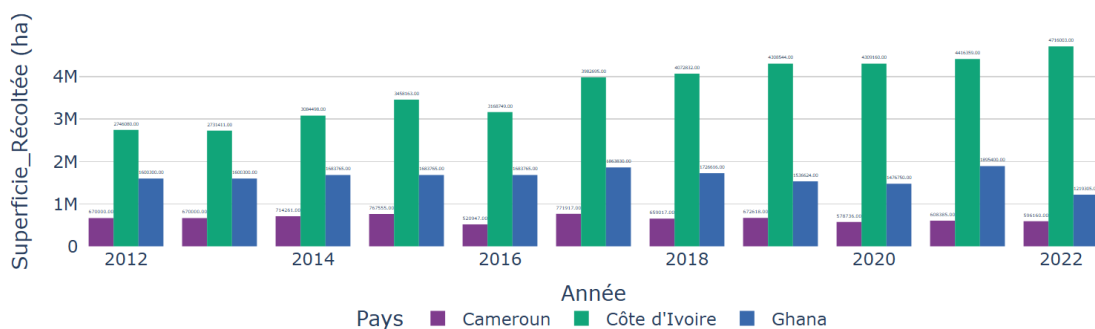
Évolution des productions de cacao (2012-2022)



10 Evolution comparée des superficies de cacao entre les trois pays (2012-2022)

La figure ci-dessous de l'évolution des superficies récoltées reflète la même hiérarchie que celle observée pour la production. La Côte d'Ivoire mobilise les plus grandes superficies, suivie du Ghana et, plus loin derrière, du Cameroun. Ici encore, les tests de sensibilité confirment la robustesse de cette hiérarchie : même en prenant en compte l'incertitude des données, l'ordre relatif reste stable. Cette situation montre que la compétitivité des deux leaders de la filière repose avant tout sur l'extension des surfaces cultivées plutôt que sur une amélioration des rendements.

Évolution des superficies de cacao (2012-2022)



11 Conclusion et recommandation

Arrivée au terme de notre analyse, les résultats sont assez contrastés. La hiérarchie en termes de production et de superficies est solide et peu affectée par la pondération des données, tandis que les comparaisons de rendement restent fragiles. Il semblerait que la croissance du cacao en Afrique subsaharienne a davantage reposé sur l'expansion des terres que sur des innovations technologiques. Ce modèle est toutefois vulnérable, car il accentue la pression sur les terres et les écosystèmes.

Concernant l'origine des données, malgré que le Cameroun et le Ghana présentent une proportion importante du caractère officielle que la Côte-d'Ivoire, il est important d'inviter ces trois pays à renforcer leurs systèmes statistiques nationaux. Le but est de réduire au maximum les imputations, les estimations et les données provenant des institutions internationales.

Pour l'avenir du cacao en Afrique subsaharienne, il est urgent de réorienter la trajectoire vers une intensification durable afin de limiter l'expansion des terres. En parallèle, il est essentiel d'améliorer la gouvernance des données : les comparaisons régionales devraient intégrer systématiquement des pondérations selon l'origine des chiffres et documenter les incertitudes associées. Enfin, les politiques publiques, qu'il s'agisse de subventions, de crédits ou de mécanismes de certification, gagneraient à conditionner leurs soutiens à l'adoption de modèles agroforestiers mesurables, afin de consolider à la fois la compétitivité économique et la durabilité environnementale de la filière.

Bibliographie

Fountain, A. C., & Hütz-Adams, F. (2022). *Cocoa Barometer 2022*. 2022-USA Edition. VOICE Network. <https://www.cocoabarometer.org>

International Cocoa Organization (ICCO). (2023). *Quarterly Bulletin of Cocoa Statistics & Monthly Cocoa Market Reports*. Abidjan : ICCO. Disponible sur : <https://www.icco.org>

Briggs, A. H., Sculpher, M., & Buxton, M. J. (1994). "Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis." *Health Policy and Planning*, 9(4), 385–394.